

## RESEARCH ARTICLE

## Open Access



# A comparison of methods to adjust for continuous covariates in the analysis of randomised trials

Brennan C. Kahan<sup>1\*</sup>, Helen Rushton<sup>2</sup>, Tim P. Morris<sup>3,4</sup> and Rhian M. Daniel<sup>4</sup>

## Abstract

**Background:** Although covariate adjustment in the analysis of randomised trials can be beneficial, adjustment for continuous covariates is complicated by the fact that the association between covariate and outcome must be specified. Misspecification of this association can lead to reduced power, and potentially incorrect conclusions regarding treatment efficacy.

**Methods:** We compared several methods of adjustment to determine which is best when the association between covariate and outcome is unknown. We assessed (a) dichotomisation or categorisation; (b) assuming a linear association with outcome; (c) using fractional polynomials with one (FP1) or two (FP2) polynomial terms; and (d) using restricted cubic splines with 3 or 5 knots. We evaluated each method using simulation and through a re-analysis of trial datasets.

**Results:** Methods which kept covariates as continuous typically had higher power than methods which used categorisation. Dichotomisation, categorisation, and assuming a linear association all led to large reductions in power when the true association was non-linear. FP2 models and restricted cubic splines with 3 or 5 knots performed best overall.

**Conclusions:** For the analysis of randomised trials we recommend (1) adjusting for continuous covariates even if their association with outcome is unknown; (2) keeping covariates as continuous; and (3) using fractional polynomials with two polynomial terms or restricted cubic splines with 3 to 5 knots when a linear association is in doubt.

**Keywords:** Randomised controlled trial, Covariate adjustment, Continuous variables, Fractional polynomials, Restricted cubic splines

## Background

Adjustment for prognostic covariates in the analysis of randomised controlled trials (RCTs) can offer substantial benefits [1–12]. These include increased power [1–6], protection against chance imbalances between treatment arms [1], and correct results when the covariate was used as a stratification factor during randomisation [1, 8–12]. Adjustment for binary or categorical covariates is relatively straightforward through the use of indicator (or dummy) variables, as there is no risk of misspecifying the nature of the association with the outcome. However, adjustment for continuous covariates is more complex, as

the shape of this association does need to be specified. For example, we could assume this association is linear, quadratic, logarithmic, or takes on some other form.

When the shape of the association between covariate and outcome is known (due to previous data, or biological or medical reasons), adjustment for continuous covariates is straightforward. However, when the association is unknown the best method of adjustment is unclear. Potential options include dichotomisation or categorisation (grouping the covariate into two or more categories); assuming a linear association between covariate and outcome; or using the data to estimate a potentially non-linear association, for example by using fractional polynomials [13] or restricted cubic splines (RCS) [14, 15].

\* Correspondence: [b.kahan@qmul.ac.uk](mailto:b.kahan@qmul.ac.uk)

<sup>1</sup>Pragmatic Clinical Trials Unit, Queen Mary University of London, London, UK  
Full list of author information is available at the end of the article

The issue of adjusting for continuous covariates has been studied in the context of observational, non-randomised studies [13, 16–19], however there has been comparatively little research into this issue in RCTs. The Committee for Proprietary Medicinal Products' (CPMP) guidance document *Points to Consider on Adjustment for Baseline Covariates* states that “... in the absence of any well-established prior knowledge about the relationship between the covariates and the outcome (which is often the case in most clinical trials) the model should use a simple form. For example, when the covariate is continuous, then the model might be based on a linear relationship between the covariate and outcome, or on a categorisation of the covariate into a few levels, the number of levels depending upon the sample size” [20]. However, these simple approaches may lead to misspecification of the association between the covariate and outcome, which can lead to a decrease in power [21]. Allowing for more complex associations between covariate and outcome may therefore be useful in order to maximise statistical power.

The goals of this paper are to investigate which methods of adjusting for a continuous covariate in the analysis of a RCT maximise power whilst still retaining correct type I error rates and unbiased estimate of treatment effect, when the true association between covariate and outcome is unknown.

## Methods

### Problems with misspecification

We begin by exploring some of the potential issues that may occur if the association between a continuous covariate and the outcome is misspecified (that is, when the assumed association is different to the true association). In general, misspecification will affect results only when there is a true association between covariate and outcome, and for the purposes of this discussion we assume that the covariate *does* influence outcome. It should be noted however that the association between covariate and outcome does not need to be causal. Finally, we only consider covariates measured before randomisation, as adjustment for post-randomisation factors can lead to biased estimates of treatment effect [22, 23].

In observational studies, one of the primary issues with misspecification is residual confounding; that is, adjusting for the misspecified covariate will not fully account for the confounding. This can lead to biased estimates and misleading conclusions. However, residual confounding is not an issue in RCTs, provided the randomisation procedure has been performed correctly, as this ensures there are no systematic differences between treatment arms (although chance imbalances can still occur) [24]. The primary concern regarding misspecification in RCTs is therefore whether it affects the general

operating characteristics of the trial, e.g. the estimate of the treatment effect, type I error rate, or power.

Current evidence suggests that misspecification of the covariate-outcome relationship will not increase the type I error rate [21, 25]. However, it can affect power and the estimated treatment effect, though these issues differ for linear vs. non-linear models (where linear models include analyses that estimate a difference in means for continuous outcomes, or a risk difference or risk ratio for binary outcomes, and non-linear models include those that estimate an odds ratio for binary outcomes, or a hazard ratio for time-to-event models with censoring).

For linear models (e.g. a difference in means or proportions), misspecification will not affect the unbiasedness of the estimator of the treatment effect; this remains unbiased regardless of the extent of the misspecification. However, the precision with which the treatment effect is estimated will be reduced, leading to a reduction in power [25].

For non-linear models (e.g. models that estimate odds ratios, or hazard ratios with censoring), misspecification will affect the estimated treatment effect; it will be attenuated towards the null, and power will be reduced [21, 26–28].

In general, the impact will depend on the extent of the misspecification (how far our assumed association is from the true association); greater degrees of misspecification will lead to greater decreases in power, and a higher degree of attenuation in treatment effect for binary or time-to-event outcomes.

### Methods of analysis

Below we outline various methods to adjust for continuous covariates, and highlight the assumptions made by each analysis. Figures 1 and 2 compare the estimated association for each method of analysis to the true association.

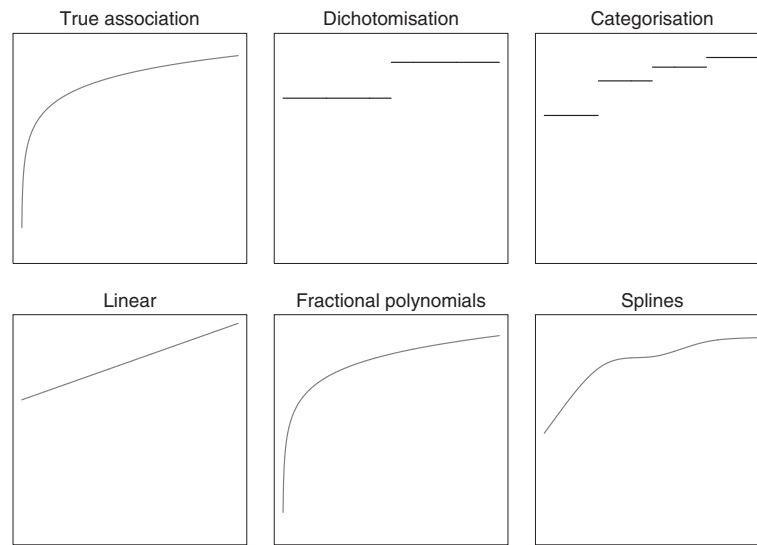
#### Dichotomisation

Dichotomisation involves splitting patients into two groups based on their covariate values. For example, patients may be grouped according to their body mass index (BMI) score as either overweight (25 or over) or not overweight (under 25). As noted by others, this type of grouping can be helpful in clinical practice, but is not necessarily helpful for data analysis [29].

A dichotomised analysis can be implemented using the following model:

$$Y = \alpha + \beta_T X_T + \beta_O X_O + \varepsilon$$

where  $X_T$  is a binary variable indicating whether the patient received treatment or control,  $\beta_T$  is the effect of receiving treatment,  $X_O$  is a binary covariate indicating



**Fig. 1** Estimated associations for different methods of analysis for  $y = \log(x)$ . \*To obtain the estimated association between  $f(x)$  and  $y$  for each method of analysis, we first generated a single data set from the model  $y = \log(x) + \epsilon$  (where  $\epsilon \sim N(0, 1)$ ). We then fit a linear regression model of the form  $y = f(x)$  to obtain the estimated association

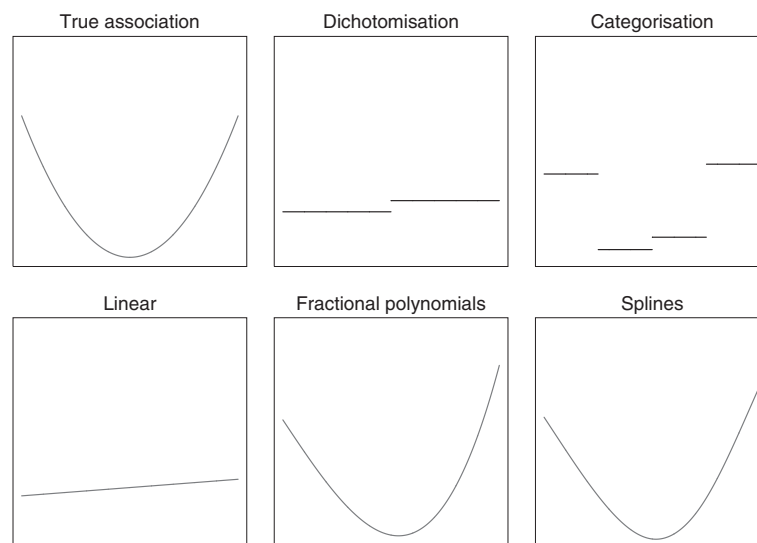
whether the patient was overweight ( $BMI > 25$ ),  $\beta_O$  is the effect of being overweight, and  $\epsilon$  is a random error term.

The primary issue with dichotomisation is that it throws away a large amount of information, which can reduce power. For example, an analysis that treated BMI as continuous would recognise that BMI values of 24 and 26 are different, but much more similar to each other than BMI values of 16 or 34 are. However, a dichotomised analysis (grouped as less or more than 25) treats values of 16 and 24 as identical, values of 26 and

34 as identical, but treats BMI values of 24 and 26 as opposite.

#### Categorisation

Categorisation is when patients are grouped into multiple categories. It is a generalisation of dichotomisation; in this paper we assume categorisation involves three or more groups, in order to distinguish it from dichotomisation.



**Fig. 2** Estimated associations for different methods of analysis for  $y = x^2$ . \*To obtain the estimated association between  $f(x)$  and  $y$  for each method of analysis, we first generated a single data set from the model  $y = x^2 + \epsilon$  (where  $\epsilon \sim N(0, 1)$ ). We then fit a linear regression model of the form  $y = f(x)$  to obtain the estimated association

Like dichotomisation, categorisation reduces the amount of information in the analysis, potentially leading to a loss in power. However, due to the increased number of categories, less information will be lost than in dichotomisation, and it should therefore give better results. For example, categorising BMI into underweight (<18.5), normal weight (18.5 to 24.9), overweight (25 to 29.9), or obese (30 or more) allows BMI scores of 16, 24, 26, and 30 to each have a different effect on outcome, unlike dichotomisation.

A categorised analysis can be implemented using the following model:

$$Y = \alpha + \beta_T X_T + \beta_N X_N + \beta_{OV} X_{OV} + \beta_{OB} X_{OB} + \varepsilon$$

where  $X_N$ ,  $X_{OV}$  and  $X_{OB}$  indicate whether the patient is normal weight, overweight, or obese respectively, and  $\beta_N$ ,  $\beta_{OV}$  and  $\beta_{OB}$  are the effects of being in these BMI categories compared to being underweight.

In general, a higher number of categories leads to less information lost [21]. However, having too many categories can be problematic, particularly with a small sample size. This can lead to reduced power due to the extra parameters in the model [1]. It can also lead to inflated type I error rates and biased estimates of treatment effect for binary or time-to-event outcomes when the number of categories is high compared to the number of events [1].

### Linear association

A linear analysis keeps the covariate as continuous, and assumes the association between covariate and outcome is linear. This analysis assumes the effect of an increase in the covariate is constant across the range of the covariate. For example, an increase in BMI from 15 to 16 would have the same impact on outcome as an increase from 29 to 30.

A linear analysis can be implemented using the following model:

$$Y = \alpha + \beta_T X_T + \beta_{BMI} X_{BMI} + \varepsilon$$

where  $X_{BMI}$  represents BMI on a continuous scale, and  $\beta_{BMI}$  represents the effect of a one-unit increase in BMI on outcome.

The primary advantage of a linear analysis over categorisation is that it makes full use of the data, and so should increase study power. However, if the true association between covariate and outcome is non-linear, then a linear analysis will be misspecified and may lead to reductions in power.

### Fractional polynomials

Fractional polynomial models use polynomial transformations to estimate the association between the

covariate and outcome. They typically use either one or two polynomial terms. A model using only one polynomial term is referred to as an 'FP1' model, and a model using two polynomials terms an 'FP2' model. An FP1 can be written as follows:

$$Y = \alpha + \beta_T X_T + \beta_1 X_{BMI}^{p_1} + \varepsilon$$

where  $p_1$  is a polynomial transformation estimated from the set  $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$  (where  $p = 0$  is taken to mean, by convention,  $\log(X)$ ).

A model with two polynomial terms (FP2) can be written as:

$$Y = \alpha + \beta_T X_T + \beta_1 X_{BMI}^{p_1} + \beta_2 X_{BMI}^{p_2} + \varepsilon$$

where  $p_1$  and  $p_2$  are polynomial transformations estimated from the set  $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$  (where  $p = 0$  corresponds to  $\log(X)$ ). By convention,  $p_1 = p_2$  is taken to mean a model in which the two terms are  $\beta_1 X_{BMI}^{p_1}$  and  $\beta_2 X_{BMI}^{p_2} \log(X_{BMI})$ .

Either FP1 or FP2 models could be used in practice. It is possible to use the trial data to select between FP1 and FP2 models, however we do not recommend this approach, as covariate selection procedures have been shown to lead to poor results [30]. We therefore recommend pre-specifying the use of either an FP1 or FP2 model, and consider both approaches in this paper. Furthermore, in some software packages fractional polynomials also incorporate a model selection algorithm, where they drop covariates which are not prognostic enough (e.g. in the example above, they may drop BMI from the final model if it does not meet some pre-defined statistical significance threshold). As mentioned earlier, we do not recommend model selection in RCTs, and so for the purposes of this paper we have only considered fractional polynomial models which include all covariates, regardless of their statistical significance.

The benefits of using a fractional polynomial approach include keeping the data as continuous, and allowing for non-linear associations. Fractional polynomials can be implemented in most standard statistical packages (e.g. using the *fp* or *mfp* commands in Stata, or the *mfp* package in R). Further details on fractional polynomials are available elsewhere [13, 31].

### Restricted Cubic Splines

RCS is implemented by splitting the continuous covariate into separate sections, separated by  $m$  different knots,  $k_1 < k_2 < \dots < k_m$ . Within each of these sections, a polynomial relationship between the covariate and outcome is estimated; these polynomial functions are joined up at the knots, to ensure a smooth curve across the range of the covariate. Two boundary knots  $k_{min} < k_1$  and  $k_{max} > k_m$  (usually placed at the extremes of the

covariate) are also used; RCS estimates a linear association between covariate and outcome in these boundary knots, i.e. between  $k_{min}$  and  $k_1$ , and  $k_m$  and  $k_{max}$ .

The model can be written as:

$$Y = \beta_T X_T + S(BMI) + \varepsilon$$

where

$$S(BMI) = \gamma_0 + \gamma_1 X_{BMI} + \sum_{j=1}^m \gamma_j \left[ (X_{BMI} - k_j)_+^3 - \lambda_j (X_{BMI} - k_{min})_+^3 - (1 - \lambda_j) (X_{BMI} - k_{max})_+^3 \right]$$

and

$$(X_{BMI} - k)_+^3 = \begin{cases} (X_{BMI} - k)^3 & \text{if } X_{BMI} \geq k \\ 0 & \text{if } X_{BMI} < k \end{cases}$$

and

$$\lambda_j = \frac{k_{max} - k_j}{k_{max} - k_{min}}$$

Although seemingly complicated, RCS can easily be implemented in most software packages (e.g. the *mkspline* command in Stata, the *effect* option in SAS, or the *hmisc* package in R). In practice one must specify the number of knots to use, and where to place them. One could estimate the optimal number and location of the knots based on the trial data, but as above, these types of model selection procedures do not always work well for RCTs, and we therefore recommend that these choices are pre-specified. In this paper, we consider both 3 and 5 knots, and placing them at specified percentiles of the data [15].

RCSs have similar benefits to fractional polynomials: they keep the data as continuous, and allow for non-linear associations. Further details on RCSs are available elsewhere [15].

### Simulation study

We performed a simulation study to compare different methods of accounting for a continuous covariate in the analysis of a RCT with both continuous and binary outcomes.

We generated outcomes from the following model:

$$Y_i = \alpha + \beta_T X_{T_i} + \beta_{cov} f(X_i) + \varepsilon_i$$

where  $X_{T_i}$  is a binary variable indicating whether the patient received treatment or control,  $\beta_T$  is the effect of receiving treatment,  $X_i$  is a continuous covariate,  $f(\cdot)$  is a transformation,  $\beta_{cov}$  is the effect of the transformed covariate, and  $\varepsilon_i$  is a random error term.

For continuous outcomes, we set  $\varepsilon_i$  to follow a normal distribution with mean 0 and standard deviation  $\sigma_e$ , with  $\sigma_e$  equal to 1.

For binary outcomes, we set  $\varepsilon_i$  to follow a logistic distribution with mean 0 and variance  $\pi^2/3$ .  $Y_i$  then represents a latent continuous outcome, and a binary response was generated as 1 if  $Y_i > 0$  and 0 otherwise. This model implies the  $\beta$ 's represent log odds ratios.

For both outcome types we assessed three scenarios for  $X$ :

- A linear association with the outcome:  $f(X) = X$
- A non-linear, monotonic association with the outcome:  $f(X) = e^X$
- A non-linear, non-monotonic association with the outcome:  $f(X) = X^2$

For each scenario we generated  $X$  from a normal distribution with mean 0 and standard deviation 1.

We chose  $\beta_{cov}$  based on the following formula:

$$\beta_{cov}(p_{90} - p_{10}) = \sigma_e$$

where  $p_z$  is the  $z^{\text{th}}$  percentile of  $f(X)$ ; that is, an increase from the 10<sup>th</sup> to the 90<sup>th</sup> percentile in  $f(X)$  would increase the outcome by one unit of  $\sigma_e$ . For continuous outcomes, this led to  $\beta_{cov}$  values of 0.385, 0.300, and 0.372 for linear, non-linear monotonic, and non-linear non-monotonic associations respectively. For binary outcomes, this led to  $\beta_{cov}$  values of 0.700, 0.550, and 0.674 for linear, non-linear monotonic, and non-linear non-monotonic associations respectively.

We set the sample size to 200 patients for continuous outcomes, and to 600 for binary outcomes. These values were selected based on a review of trials published in high impact general medical journals which found these were the median sample sizes for trials with each outcome type [8]. Patients were randomised to one of two treatment arms using simple randomisation. For each simulation scenario (linear, monotonic, non-monotonic) we used two treatment effects:  $\beta_T$  was set to 0, or  $\beta_T$  was set to give 80 % power based on the sample size (assuming correct specification of the association between covariate and outcome). For binary outcomes we set the event rate in the control arm to 50 %.  $\beta_T$  was set to give 80 % power based on both the sample size and the effect of  $\beta_{cov}$  on outcome; because the effect of  $\beta_{cov}$  differed according to the scenario, this implies that  $\beta_T$  was set to different values depending on the type of association between covariate and outcome.

We analysed continuous outcomes using a linear regression model, and binary outcomes using a logistic regression model. We adjusted for the continuous covariate  $X$  in the regression model using seven different approaches: (a) dichotomising  $X$  at its sample median; (b) categorising  $X$  at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> sample percentiles; (c) including  $X$  as a continuous covariate,



assuming a linear association; (d) using fractional polynomials, with one polynomial term (FP1); (e) using fractional polynomials, with two polynomial terms (FP2); (f) using restricted cubic splines with 3 knots (knots were placed based on Harrell's recommended percentiles [15]); and (g) using restricted cubic splines with 5 knots (knots were placed based on Harrell's recommended percentiles [15]).

For each scenario we calculated the bias in the estimated treatment effect, the type I error rate (when  $\beta_T = 0$ ) and the power (when  $\beta_T \neq 0$ ). For each simulation scenario we used 5000 replications.

### Re-analysis of MIST2 and APC trials

We applied the different methods of accounting for continuous covariates to the MIST2 and advanced prostate cancer (APC) trials. MIST2 compared four treatments for patients with pleural infection [32]; placebo, tPA, DNase, or tPA + DNase. We focus on the treatment comparison between tPA + DNase vs. placebo for simplicity. We used a logistic regression model to re-analyse the outcome of surgery at three months. Of the 192 patients included in the analysis, 31 (16 %) experienced an event. We adjusted for the size of the patient's pleural effusion at baseline (continuous covariate), as well as two binary covariates: whether the infection was hospital acquired and whether the infection was purulent. All three covariates were minimisation factors. In our re-analysis, we handled the continuous covariate (size of the patient's pleural effusion) in eight different ways: (a) we excluded it; (b) we dichotomised it at its sample median; (c) we categorised it at its sample 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles; (d) we included it as a continuous covariate, assuming a linear association with outcome; (e) we used an FP1 model; (f) we used an FP2 model; (g) we used

RCS with 3 knots; and (h) we used RCS with 5 knots. For the fractional polynomial models, we forced the model to include the covariate regardless of its significance level, and for the RCS models, we placed the knots at the percentiles recommended by Harrell.

The APC trial compared diethyl stilboestrol vs. placebo on overall survival in patients with advanced prostate cancer. We used a Cox regression model to re-analyse the outcome of overall mortality. In our re-analysis we used the dataset supplied by Royston and Sauerbrei [13]. Of 475 patients included in the analysis, 338 (71 %) experienced an event. We adjusted for three continuous covariates: patient weight, tumour size, and stage grade. All three are prognostic factors associated with mortality. We used the same methods of analysis as for the MIST2 trial above. We analysed each of the three continuous covariates using the same method (that is, dichotomised all three, used an FP2 model for all three, etc.).

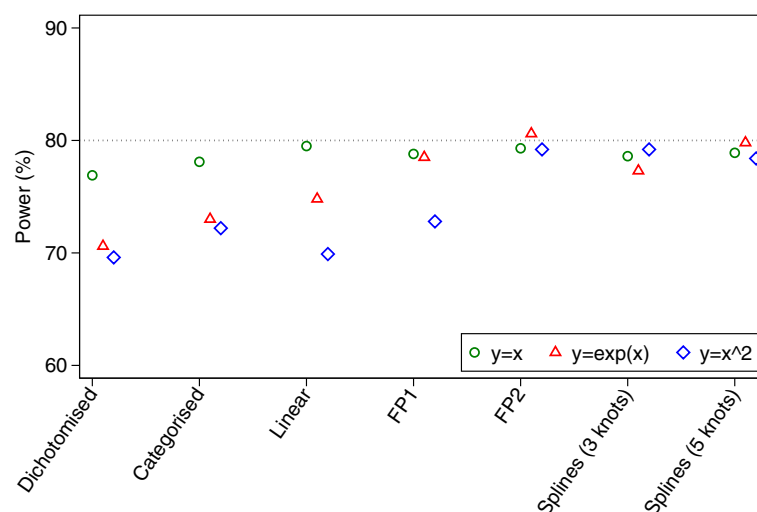
## Results

### Simulation results for continuous outcomes

All methods of analysis provided unbiased estimates of treatment effect, and correct type I error rates (range 4.7 to 5.7 %) across all scenarios.

Results for power are shown in Fig. 3. For data generated under a linear association, all methods of analysis which kept the covariate  $X$  as continuous (linear analysis, fractional polynomials, and restricted cubic splines) had nominal power. Conversely, dichotomisation and categorisation led to small reductions in power.

For data generated under a non-linear, monotonic association, FP and splines gave the highest power. A linear analysis led to a reduction in power of about 5.8 % compared to FP2, and dichotomisation and categorisation gave reductions of about 10.0 and 7.6 % respectively.



**Fig. 3** Simulation results for continuous outcomes (power)

For data generated under a non-linear, non-monotonic association, FP2 and splines with 3 or 5 knots gave the highest power. A linear analysis and dichotomisation had the lowest power (9.3 and 9.6 % reductions respectively vs. FP2), and categorisation lost 7.0 % power. The FP1 model lost 6.4 % power compared to FP2.

### Simulation results for binary outcomes

Results are shown in Figs. 4 and 5. All methods of analysis were unbiased when the treatment had no effect (i.e. when the odds ratio = 1), and gave correct type I error rates (range 4.6 to 5.7 %) across all scenarios. When the treatment was effective (i.e. when the odds ratio  $\neq 1$ ), dichotomisation, categorisation, a linear analysis, and FP1 models all led to bias in certain scenarios. This was most pronounced for data generated under a non-linear, non-monotonic association; dichotomisation, a linear analysis, and FP1 all led to the log(OR) being biased downwards by around 20 %. Conversely, FP2 and splines with 3 or 5 knots all produced unbiased estimates across all scenarios. Results are shown in Fig. 4.

For data generated under a linear association, a linear analysis, FP, and splines all gave unbiased estimates of treatment effect and similar power. The log odds ratios for dichotomisation and categorisation were attenuated by 11.7 and 3.0 % respectively, leading to a reduction in power of 5.2 and 1.6 % compared to a linear analysis.

Under a non-linear monotonic association, a linear analysis, FP, and splines all gave unbiased estimates of treatment effect and nominal power. Dichotomisation and categorisation had very little attenuation in the estimated treatment effects, and small reductions in power compared to FP (1.3 % dichotomisation, 0.5 % categorisation).

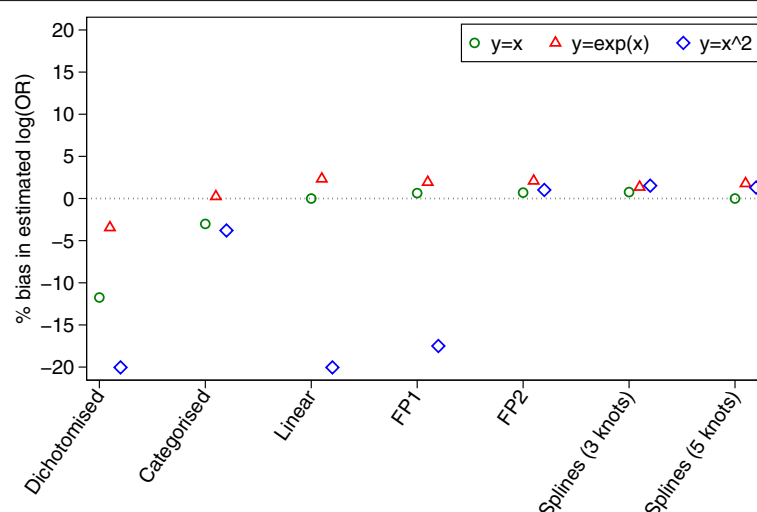
Under a non-linear, non-monotonic association, FP2 and splines with 3 or 5 knots gave unbiased treatment estimates and good power. Categorisation led to a small attenuation of the estimated log odds ratio, leading to a small decrease in power compared to FP2. A linear analysis and dichotomisation both led to substantially attenuated treatment effects, leading to large decreases in power compared to FP2 (9.1 % linear, 9.3 % dichotomisation). The FP1 model also led to a large degree of bias in the estimated treatment effect, and subsequently a large reduction in power (8.6 %) compared to FP2. This is because FP1 only allows for a monotone association between  $X$  and  $Y$ .

### Results of re-analysis of MIST2 and APC trials

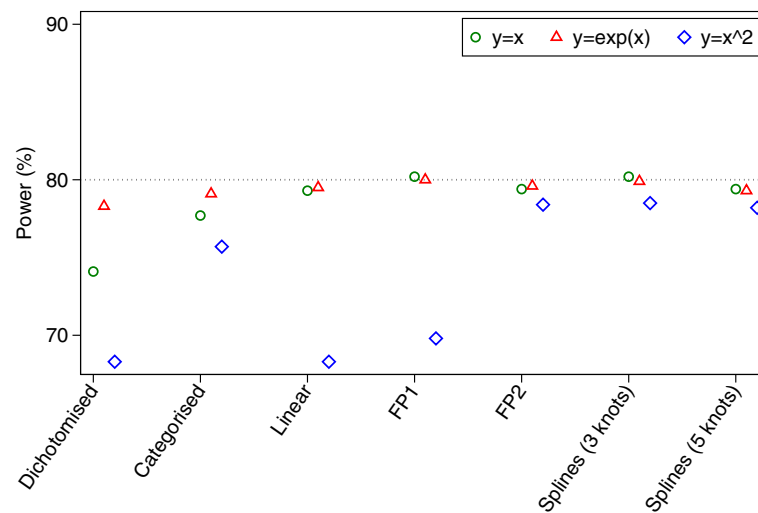
Results can be found in Table 1. In both trials, unadjusted/dichotomised/categorised analyses led to smaller treatment effect estimates than linear/FP/spline analyses. Treatment effect estimates were reduced by 35, 24, and 6 % for unadjusted, dichotomised, and categorised analyses respectively, compared to FP or splines in MIST2. This attenuation in the estimated treatment effects led to larger p-values for unadjusted, dichotomised, and categorised analyses in most cases, which sometimes led to results becoming non-significant.

### Discussion

Misspecification of the association between a continuous covariate and the outcome in RCTs can lead to substantial reductions in power. This occurs due to a reduction in the precision of the estimated treatment effect for linear analyses (such as continuous outcomes, or a difference in proportions) and a reduction in the size of the estimated treatment effect for non-linear models (such as a binary or time-to-event outcome with censoring



**Fig. 4** Simulation results for binary outcomes (bias)



**Fig. 5** Simulation results for binary outcomes (power)

estimated using an odds or hazard ratio). The extent to which results are affected is influenced by the extent of the misspecification.

Re-analysis of the MIST2 and APC trials found that omitting a covariate from the analysis led to larger attenuation of the estimated treatment effect compared to including the covariate, even if the association between covariate and outcome was misspecified. This is likely because excluding a covariate from the analysis can be seen as a more severe form of misspecification, and therefore resulted in larger losses in precision and attenuation of treatment effects. Therefore, we recommend adjusting for covariates even if the true association is unknown.

Our simulation study demonstrated that analyses which keep covariates as continuous generally perform better than analyses that use dichotomisation or categorisation. The simplest method of keeping a covariate continuous is to assume a linear association with the outcome. A linear analysis will perform well if the association between covariate and outcome is approximately

linear. However, there may be large reductions in power in the presence of departures from linearity. If non-linearity is possible, then FP2 models or splines are both suitable options, as they have been shown to increase power compared to alternative methods. FP1 models should be used with caution, as these provide very poor results under non-linear, non-monotonic associations. This is because FP requires all covariates to take only positive values, and therefore the outcome is a monotone function of the covariates for any FP1 model.

One issue to consider when deciding if a covariate is likely to have a linear association with outcome is the expected range of the covariate within the trial population. This range will often be smaller in trials than in observational studies due to more restrictive inclusion/exclusion criteria. Covariates with a non-linear association across their entire range may actually have a linear or approximately linear association across certain subsets of their range. For example, imagine BMI is non-linear across the range 16–35. Then, for any small portion of this range (e.g. 16–20, 20–24, etc.), the

**Table 1** Results from different methods of adjusting for continuous covariates in APC and MIST2 trials

Analysis methods	APC		MIST2	
	HR (95 % CI)	P-value	OR (95 % CI)	P-value
Unadjusted	0.84 (0.68 to 1.04)	0.11	0.23 (0.05 to 1.15)	0.07
Dichotomised	0.82 (0.66 to 1.01)	0.06	0.21 (0.04 to 1.07)	0.06
Categorised	0.82 (0.66 to 1.01)	0.07	0.18 (0.03 to 0.94)	0.04
Linear	0.79 (0.64 to 0.98)	0.03	0.16 (0.03 to 0.87)	0.03
FP1	0.79 (0.64 to 0.98)	0.03	0.17 (0.03 to 0.90)	0.04
FP2	0.78 (0.63 to 0.97)	0.03	0.17 (0.03 to 0.90)	0.04
Splines (3 knots)	0.80 (0.64 to 0.99)	0.04	0.18 (0.03 to 0.96)	0.04
Splines (5 knots)	0.81 (0.65 to 1.00)	0.06	0.17 (0.03 to 0.91)	0.04



association may be at least approximately linear. Therefore, in a trial recruiting only overweight patients (BMI 25–29.5), a linear analysis may be appropriate.

Continuous covariates are often categorised when used as stratification factors during randomisation. Stratification or minimisation induces correlation between treatment arms, and it is therefore necessary to account for this correlation in the analysis to obtain valid standard errors and type I error rates [8, 9]. In practice, this means we must include the stratification factors in our analysis. It is therefore of interest to know whether we must use the (categorised) stratification factor, or whether we can use the continuous version. Theoretically, correctly modelling the functional form of the continuous covariate in the analysis should adequately account for the correlation induced by the stratified randomisation procedure, and so this approach should lead to valid standard errors and type I error rates, as well as increased power. However, further research to confirm this hypothesis would be useful.

Both fractional polynomials and restricted cubic splines require certain decisions to be made about their implementation (e.g. FP1 or FP2, number and placement of knots, etc.), and in practice, we could use the trial data to make these choices. For example, one could estimate the optimal number and location of the knots. However, when using trial data to select the model, there is a risk of model overfitting, which can lead to poor results. We therefore suggest that model selection be kept to a minimum, and that the form of the covariates be pre-specified [33, 34]. In addition to pre-specifying the general analysis approach (e.g. assuming a linear association vs. fractional polynomials vs. restricted cubic splines), it is necessary to pre-specify the implementation of these approaches. For fractional polynomials, this entails pre-specifying the whether an FP1 or FP2 model will be used. For restricted cubic splines, this entails pre-specifying the number and location of the knots. We also note that both fractional polynomials and RCS can be combined with model selection algorithms which determine which covariates should be kept in the final model, and which covariates should be discarded (usually based on a statistical significance threshold). However, analysis methods which rely on p-values to determine the form of the final model have been shown to give poor results in RCTs in a variety of scenarios [30, 35–37], and so we do not recommend this approach. Instead, we recommend when using fractional polynomials or splines that all covariates are included in the model regardless of their statistical significance.

## Conclusion

We recommend (1) adjusting for continuous covariates even if their association with outcome is unknown; (2)

keeping covariates as continuous; and (3) using fractional polynomials with two polynomial terms or restricted cubic splines with between 3–5 knots when a linear association is in doubt.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data and materials

Data from the APC trial are available in Royston P, Sauerbrei W. *Multivariable Model-Building*. Chichester, England. : Wiley; 2008. The authors of this manuscript do not have permission to share data from the MIST2 trial.

## Abbreviations

APC: Advanced prostate cancer; BMI: Body mass index; CPMP: Committee for Proprietary Medicinal Products; FP: Fractional polynomial; MIST2: The second multicentre intrapleural sepsis trial; RCT: Randomised controlled trial; RCS: Restricted cubic splines.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

BK: concept, design of simulation study, conducted simulations, wrote manuscript. HR: design of simulation study, conducted simulations, contributed to manuscript. TM: design of simulation study, contributed to manuscript. RD: design of simulation study, contributed to manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank the reviewer for helpful comments on the manuscript.

## Funding

No authors received specific funding for this work. Tim Morris is funded by the MRC London Hub for Trials Methodology Research, grant MC\_EX\_G0800814.

## Author details

<sup>1</sup>Pragmatic Clinical Trials Unit, Queen Mary University of London, London, UK. <sup>2</sup>University of Manchester, Manchester, UK. <sup>3</sup>MRC Clinical Trials Unit at UCL, London School of Hygiene and Tropical Medicine, London, UK. <sup>4</sup>Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, UK.

Received: 9 September 2015 Accepted: 1 April 2016

Published online: 11 April 2016

## References

1. Kahan BC, Jairath V, Dore CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014;15:139.
2. Hernandez AV, Eijkemans MJ, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? *Ann Epidemiol*. 2006;16(1):41–8.
3. Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol*. 2004;57(5):454–60.
4. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21(19):2917–30.

5. Turner EL, Perel P, Clayton T, Edwards P, Hernandez AV, Roberts I, et al. Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. *J Clin Epidemiol*. 2012;65(5):474–81.
6. Thompson DD, Lingsma HF, Whiteley WN, Murray GD, Steyerberg EW. Covariate adjustment had similar benefits in small and large randomized controlled trials. *J Clin Epidemiol*. 2015;68(9):1068–75.
7. Nicholas K, Yeatts SD, Zhao W, Ciolino J, Borg K, Durkalski V. The impact of covariate adjustment at randomization and analysis for binary outcomes: understanding differences between superiority and noninferiority trials. *Stat Med*. 2015;34(11):1834–40.
8. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*. 2012;345:e5840.
9. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med*. 2012;31(4):328–40.
10. Kahan BC, Morris TP. Adjusting for multiple prognostic factors in the analysis of randomised trials. *BMC Med Res Methodol*. 2013;13:99.
11. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. *BMC Med Res Methodol*. 2013;13:58.
12. Parzen M, Lipsitz S, Dear K. Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? *Biom J*. 1998;40(4):385–402.
13. Royston P, Sauerbrei W. *Multivariable Model-Building*. Chichester: Wiley; 2008.
14. Durrleman S, Simon R. Flexible regression models with cubic splines. *Stat Med*. 1989;8(5):551–61.
15. Harrell Jr FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
16. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology (Cambridge, Mass)*. 1997;8(4):429–34.
17. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002;7(1):19–40.
18. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127–41.
19. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26(30):5512–28.
20. Committee for Proprietary Medicinal Products (CPMP). Points to consider on adjustment for baseline covariates. *Stat Med*. 2004;23(5):701–9.
21. Schmoor C, Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Stat Med*. 1997;16(1–3):225–37.
22. Berger VW. Valid adjustment of randomized comparisons for binary covariates. *Biom J*. 2004;46(5):589–94.
23. Rosenbaum PR. The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment. *J Roy Stat Soc A Sta*. 1984;147:656–66.
24. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials*. New York: John Wiley & Sons, Inc.; 2005.
25. Yang L, Tsiatis A. Efficiency Study of Estimators for a Treatment Effect in a Pretest-Posttest Trial. *The American Statistician*. 2001;55(4):314–21.
26. Gail M, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71(3):431–44.
27. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials*. 1998;19(3):249–56.
28. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *Int Stat Rev*. 1991;58:227–40.
29. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332(7549):1080.
30. Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. *Control Clin Trials*. 2000;21(4):330–42.
31. Morris TP, White IR, Carpenter JR, Stanworth SJ, Royston P. Combining fractional polynomial model building with multiple imputation. *Stat Med*. 2015;34(25):3298–317.
32. Rahman NM, Maskell NA, West A, Teoh R, Arnold A, Mackinlay C, et al. Intrapleural use of tissue plasminogen activator and DNase in pleural infection. *N Engl J Med*. 2011;365(6):518–26.
33. Kahan BC, Jairath V, Murphy MF, Dore CJ. Update on the transfusion in gastrointestinal bleeding (TRIGGER) trial: statistical analysis plan for a cluster-randomised feasibility trial. *Trials*. 2013;14:206.
34. Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
35. Freeman PR. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Stat Med*. 1989;8(12):1421–32.
36. Kahan BC. Bias in randomised factorial trials. *Stat Med*. 2013;32(26):4540–9.
37. Shuster JJ. Diagnostics for assumptions in moderate to large simple clinical trials: do they really help? *Stat Med*. 2005;24(16):2431–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

